

Chapter 3

Test Development Process

The 1998-99 assessment cycle did not follow the usual test development timelines. Because the Kentucky Department of Education chose to re-bid the contract through a request for proposal, pretest development and item selection for the 1999 Kentucky Core Content Tests (KCCT) did not occur until October, 1998. The test development process outlined in this chapter provides a description of the procedures followed. It is recognized that this does not represent the desired test development process, but the compressed timelines necessitated these streamlined procedures.

Content Advisory Committees

The item development and selection process is based on the work of the Content Advisory Committees (CAC). The members of these committees are appointed by the Kentucky Department of Education. Each grade and subject included in the Kentucky Core Content Tests are represented by a Content Advisory Committee. Each of these committees of ten involved classroom teachers, school administrators, and university personnel. These representatives are drawn from throughout the service center regions to ensure geographic balance. Similarly, efforts are made to provide ethnic representation for each grade and subject committee.

The Test Development Sequence

As mentioned previously, the item selection and pretest development occurred simultaneously at the Content Advisory Committee meetings in October 1998. The table below outlines the steps carried out in the selection and development process.

Table 3.1
MAJOR STEPS IN THE TEST DEVELOPMENT PROCESS

1.	Analyze score data from previous year's test.
2.	Review data from previous year's test to determine which items do not meet statistical guidelines for inclusion in the KCCT.
3.	Select items for statistical linking from previous year's matrix items.
4.	Review and revise scoring guides for selected open-response items.
5.	Review content codes for selected items to ensure match to Core Content.
6.	Compare content coverage for selected items to test blueprint to identify test development needs.
7.	Develop preliminary pretest items.
8.	Draft preliminary scoring guides for pretest open-response items.
9.	Bias review of pretest items.
10.	Editorial review and revision of matrix and pretest items.
11.	Review of matrix item selections.
12.	Review of drafted pretest items.
13.	Final pretest item selection.
14.	Assemble test forms.
15.	Review test forms for balance and cueing effects.
16.	Develop camera-ready final copy.

1. ANALYZE SCORE DATA FROM PREVIOUS YEAR'S TEST.

The statistical performance of an item in the previous year's assessment was central to the item selection process. Item analyses provided information about how well items performed. This information was used to determine which items from the previous year's assessment could be considered for inclusion in the 1999 Kentucky Core Content Tests.

As mentioned in Chapter 2, there were no clear statistical guidelines provided for the selection of open-response items. Rather, the analysis and review of open-response data involved comparative judgments. The Content Advisory Committee members reviewed the data for the open-response items and compared the item means and distributions across the set of items. Particular attention was paid to those items that produced a disproportionate number of blanks, zeros, or ones.

Statistical guidelines were established for the selection of multiple-choice items. These included:

- Item mean between .40 and .80
- Biserial correlation for the correct answer .25 or above.
- Negative biserial correlations for incorrect response options.

2. REVIEW DATA FROM PREVIOUS YEAR'S TEST TO DETERMINE WHICH ITEMS (MATRIX AND PRETEST) DO NOT MEET STATISTICAL GUIDELINES FOR INCLUSION IN THE KENTUCKY CORE CONTENT TEST.

WestEd was provided the above general statistical guidelines by the Kentucky Department of Education to use in item selection. Not all items selected fell within the identified parameters. For some subject areas, particularly high school math and science, both multiple-choice and open-response items were found to be more difficult than desired. For those items, the Content Advisory Committees were asked to review the items for Core Content and grade-level appropriateness. If the item was seen as appropriate, the question was retained in the matrix. Where possible, items were edited to provide clarification to improve item functioning by changing multiple-choice distractors that were found to draw too many students or by clarifying wording in open-response items as suggested by the review of student responses.

In order to meet the new test design of the Kentucky Core Content Tests, all 1998 KIRIS matrix and pretest items were needed for the subject areas of reading, math, science, and social studies. Not all pretest items were found to function as expected, and significant edits were required. WestEd, in consultation with KDE, made case-by-case decisions about the degree of edits required for an item. If the edit was viewed as too significant, the question was not used as a matrix item. This required that some items be repeated across forms.

The Content Advisory Committees were trained on the interpretation of item statistics, the statistical guidelines to be used in item selection, the test blueprint, and the change in the test design and its impact on the number of items needed for form construction.

3. SELECT ITEMS FOR STATISTICAL LINKING FROM PREVIOUS YEAR'S MATRIX ITEMS.

In order to provide information about changes in student performance from year to year, a linking item design was used. The Content Advisory Committees reviewed the item statistics for both open response and multiple choice to select those items that reflected the desired item characteristics and represented a range of core content. These items were selected and used without revision.

4. REVISE SCORING GUIDES FOR SELECTED OPEN-RESPONSE ITEMS.

Once open-response items were selected and edited as needed, the Content Advisory Committees reviewed the scoring guides to ensure that the expectations of student performance called for in the scoring guide were appropriate for the grade level and the item as worded. Revisions were made as appropriate. No changes were made to the scoring guides for linking items to ensure the same scoring standard was applied to these items across years.

5. REVIEW CONTENT CODES FOR SELECTED ITEMS TO ENSURE MATCH TO CORE CONTENT.

The Content Advisory Committees also reviewed each item to ensure that the core content and academic expectations addressed by the item were appropriately identified. These codes were confirmed for each item selected.

6. COMPARE CONTENT COVERAGE FOR SELECTED ITEMS TO TEST BLUEPRINT TO IDENTIFY TEST DEVELOPMENT NEEDS.

Once the items were selected and the content codes were confirmed, WestEd consultants led the Content Advisory Committees through a review of the breadth of content coverage of the selected items compared to the test blueprint. This process helped to identify those areas in which the blueprint was met as well as those areas in need of additional item development.

7. DEVELOP PRELIMINARY PRETEST ITEMS.

The Content Advisory Committees were responsible for developing initial drafts of items to be pretested. Committee members were given writing assignments based on areas of need as identified by the comparison to the test blueprint. They were asked to draft both multiple-choice and open-response items. As part of the item development process, the full committee reviewed all items in their draft form. The committees were asked to evaluate for item match to the core content and grade-level appropriateness. Recommendations for item changes were incorporated by WestEd staff.

8. DRAFT PRELIMINARY SCORING GUIDES FOR PRETEST OPEN-RESPONSE ITEMS.

The Content Advisory Committees also drafted preliminary scoring guides as the open-response items were developed. This parallel development was important to ensure that the expectations for students as defined by the scoring guide were clearly reflected in the item. Committee members were asked to draft the expectations for student performance in a level 4 response as well as examples of what students would be expected to write in response to the

item. Just as the items were reviewed for match to core content and grade-level appropriateness, the scoring guides received the same scrutiny.

9. BIAS REVIEW OF PRETEST ITEMS.

All drafted pretest items were submitted to the Bias Review Committee for consideration. The role of this committee was to ensure that the content of the items was fair and equitable for all students, and that the items did not contain material that could be considered racially-ethnically-, regionally-, economically-, or gender-stereotyped or biased toward any group. Committee members used the *Guidelines for Handling Sensitive Issues in Kentucky's State Assessment Development* to guide their considerations.

The committee evaluated each item to determine if it was acceptable as submitted, acceptable with revision, or rejected. The judgments of this committee were reviewed as advisory. If after consultation with KDE staff an item was deemed acceptable with revision, an item that had been rejected by the Bias Review Committee could be used.

10. EDITORIAL REVIEW AND REVISION OF MATRIX AND PRETEST ITEMS.

WestEd's test development staff was responsible for editing the draft items. Working from the initial drafts developed by the Content Advisory Committee and incorporating the feedback of the Bias Review Committee, the development staff reviewed and revised all items considered for inclusion in the assessment. This review focused on editorial and format features of the items to ensure that they met the basic guidelines for format. Once the development staff finished their work, an item editor reviewed all pretest items.

11. REVIEW OF MATRIX ITEM SELECTIONS.

KDE staff met with WestEd test development staff for a face-to-face review of test items. For matrix items, this review consisted of an evaluation of the overall content coverage of the matrix items and an analysis of those items that had received the most substantive edits. This review determined whether those items that had been edited significantly could be used or were to be dropped from the assessment.

12. REVIEW OF DRAFTED PRETEST ITEMS.

All proposed pretest items were reviewed by KDE and WestEd staff as part of the face-to-face meetings. Items were reviewed editorially as well as for match to core content and grade-level appropriateness.

13. FINAL PRETEST ITEM SELECTIONS.

Once all the pretest items had been reviewed, KDE and WestEd staff determined the items that were to be included for pretest purposes. These selections were based on item quality and content coverage needs. As part of the selection process, pretest scoring guides were reviewed with KDE staff. Edits to both items and scoring guides were completed as agreed.

14. ASSEMBLE TEST FORMS.

Following all item selections, WestEd test development staff assembled the test forms. The balance of content coverage was the primary concern when constructing test forms.

Following content coverage, items were assigned to balance item difficulty across forms. The visual complexity of the forms was addressed by attempting to balance the number of graphics across forms. Where necessary, response options were reordered so that no clear pattern emerged in the correct answer choices. Item assignments also were affected by the designation of Form 1 for visually impaired students and Form 2 for hearing impaired students. When assigning items to the visually impaired form, special attention was given to the type of graphics included. Simple graphics and maps were included. Items that required students to evaluate complex visual stimuli (as in Arts and Humanities) were avoided. For the hearing-impaired form, items that required students to be familiar with specific sounds were avoided, but students could be asked about the principles or dynamics of sound (as in Science and Arts and Humanities).

15. REVIEW TEST FORMS FOR BALANCE AND CUEING EFFECTS.

Once the forms were assembled, each was reviewed to determine if there was an appropriate mixture of core content and academic expectations. In addition, items both within and across subject areas were reviewed for possible cueing of correct answers. Items were reassigned to forms as needed.

16. DEVELOP CAMERA-READY COPY.

Following a series of editorial and proofing checks, WestEd sent camera-ready copy to Data Recognition Corporation (DRC) for printing. DRC also provided an editorial review of all test forms.

Development of Scoring Guides

WestEd and Data Recognition Corporation (DRC) shared responsibility for the development of scorer training materials. WestEd revised the scoring rubrics for the matrix open-response items based on the items' performance in the previous year and the recommendations of the Content Advisory Committee members. The initial drafts of pretest scoring guides were developed in conjunction with the members of Content Advisory Committee and were further developed by WestEd test development staff. KDE reviewed the pretest guides as part of the face-to-face item review.

WestEd staff worked collaboratively with DRC scoring directors to identify anchor, training, and qualifying papers to be used in the training of scoring staff. Where necessary, the scoring guides were edited to provide a clear match of the anchors to the language of the scoring guide. Multiple anchors were identified for each score point. Where there was more than one way to earn a given score point, an anchor was identified for each score option. Following the identification of anchors, training papers were selected. These papers served to both reinforce the anchors and provide scorers exposure to responses that were more problematic in that they did not clearly match the language of the scoring guide. Qualifying papers were used to evaluate whether scorers were appropriately interpreting the scoring guide and accurately scoring sample papers before scoring live student work.